

# SNS 특징정보를 활용한 마르코프 논리 네트워크 기반의 단문 텍스트 분류 방법

이은지<sup>†</sup>, 김판구<sup>\*\*</sup>

## A Method for Short Text Classification using SNS Feature Information based on Markov Logic Networks

Eunji Lee<sup>†</sup>, Pankoo Kim<sup>\*\*</sup>

### ABSTRACT

As smart devices and social network services (SNSs) become increasingly pervasive, individuals produce large amounts of data in real time. Accordingly, studies on unstructured data analysis are actively being conducted to solve the resultant problem of information overload and to facilitate effective data processing. Many such studies are conducted for filtering inappropriate information. In this paper, a feature-weighting method considering SNS-message features is proposed for the classification of short text messages generated on SNSs, using Markov logic networks for category inference. The performance of the proposed method is verified through a comparison with an existing frequency-based classification methods.

**Key words:** Short Text Classification, Markov Logic Networks, Feature Weighting

### 1. 서 론

현재 스마트폰의 보급과 더불어 짧은 메시지를 이용해 여러 사람과 소통할 수 있는 소셜 네트워크 서비스(Social Networks Service)가 확산되었으며, 사용자는 정보생산의 주체가 되어 개인의 관심사와 정보를 공유 할 수 있게 되었다. 하지만, 이러한 SNS의 확산은 무분별한 연결과 정보의 과부하 문제[1]까지 초래하게 되었다. 이러한 문제를 극복하기 위해 정보 처리 기술에 대한 관심이 증가[2]하고 있으며, 그 중 문서 자동 분류[3]는 주어진 문서를 미리 정의된 카테고리에 자동으로 분류하는 기술로써, 불필요한 정

보들을 필터링하기 위해 활용되고 있다. 문서 자동 분류의 일반적인 구조는 학습데이터셋을 구축하는 단계와 분류 단계로 구분된다[4]. 첫 번째 학습데이터셋 구축단계는 각 카테고리를 서술할 수 있는 모델을 만드는 과정이며, 학습데이터셋을 기반으로 분류를 수행하기 때문에, 학습데이터가 각 카테고리의 특징을 얼마나 잘 서술하는지가 성능에 많은 영향을 준다. 기존 문서 분류에서는 문서에 등장하는 단어 간의 의미적 또는 통계적 정보를 특징으로 이용하여 학습데이터셋을 구축하지만, SNS 메시지의 경우 단문텍스트이기 때문에 문맥정보를 활용하기 쉽지 않다. 따라서 본문에서는 SNS 메시지의 특징을 고려한

※ Corresponding Author : Pankoo Kim, Address: 309, Pilmundae-ro, Dong-gu, Gwangju, Korea, TEL : +82-62-230-7799, E-mail : pkkim@chosun.ac.kr  
Receipt date : Mar. 15, 2017, Revision date : June 1, 2017  
Approval date : June 9, 2017

<sup>†</sup> Dept. of Computer Engineering, Chosun University  
(E-mail : eunbesu@gmail.com)

<sup>\*\*</sup> Dept. of Computer Engineering, Chosun University  
※ This work was supported by the Human Resource Training Program for Regional Innovation and Creativity through the Ministry of Education and National Research Foundation of Korea(NRF-2014H1C1A1073115) and the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (No. NRF-2016R1A2B4012638).

특징 가중치 부여방법을 제안하고자한다. 두 번째 분류 단계는 미분류 문서에 대한 분류를 수행하는 과정이다. 문서 분류에 적용되는 대표적인 알고리즘은 베이저안 네트워크 모델, SVM(Support Vector Machine), KNN(K-Nearest Neighbors) 등[3]이 있다. 그 중 나이브 베이저안 모델은 단순한 모델임에 비해, 성능이 높기 때문에 많이 활용 되고 있지만, 모델의 순환 관계를 형성할 수 없어 다양한 특징(Feature)에 대한 추론이 어렵다는 단점을 가지고 있다. 하지만 문서 분류의 성능 향상을 위해 다양한 특징이 선정 되어야 함에 따라 본 논문에서는 불확실성에 대한 확률이론으로 특징들의 순환 관계를 표현할 수 있는 마르코프 네트워크 모델을 문서분류에 이용하고자 한다. 논문의 구성은 다음과 같다. 2장에서는 자동문서분류에 관한 기존 연구와 마르코프 논리 네트워크에 대해 소개하고, 3장에서는 본 논문에서 제안하는 문서 분류 방법에 대해 상세히 기술한다. 4장에서는 제안한 방법을 적용할 실험 및 비교 평가를 하며, 5장에서는 결론 및 향후 연구 방향을 제시한다.

## 2. 관련 연구

### 2.1 문서 분류

문서 분류는 문서 및 문서의 카테고리를 대표할 수 있는 특징들을 이용해 새로운 문서를 해당 범주로 자동 분류하는 기법이다[5]. 문서 분류에 대한 연구는 크게 특징 선택(Feature Selection)에 대한 연구와 분류 모델 학습 알고리즘에 대한 연구로 나누어져 수행 되고 있다. 특징 선택은 문서 내 출현단어들 중 문서 분류에 적합한 특징들을 선정함으로써 효과적인 분류 모델을 생성하기 위한 연구로써 대표적인 방법인 TF-IDF[6]은 단어의 중요도는 문서 내에서 해당 단어가 많이 출현할 수 록 증가하며, 전체 자료 내에서 해당 단어가 많이 나타날수록 감소한다는 이론을 나타내는 통계적 수치이다. 상호정보량(Mutual Information)은 두 단어 중 한 단어가 다른 단어에 대해 갖고 있는 정보량을 나타내는 수치로써 문서에서 특정 단어의 출현이 문서의 범주에 포함되는지 여부를 예측하는데 사용한다[7]. 문서 분류 모델 학습에 사용되는 대표적인 알고리즘은 베이저안 모델, SVM(Support Vector Machine), KNN(K-Nearest Neighbors), C4.5 등[5]이 존재하며, 각 알고리즘마

다 선형, 규칙, 확률 기반 등 다른 분류 특징을 가지고 있기 때문에 문서에서 나타나는 각 특징들의 차이를 고려하여 학습 알고리즘을 적용하는 것이 문서의 분류의 정확도에 많은 영향을 미친다[8]. 이에 [5]의 연구에서는 웹 문서의 장르 분류에 제안한 특징 선택기법이 최적화 될 수 있는 학습알고리즘을 선정하기 위해 학습데이터에 여러 분류알고리즘을 적용함으로써, 가장 적합한 학습 알고리즘 선정을 통해 문서 분류를 수행하였다.

### 2.2 마르코프 논리 네트워크

마르코프 네트워크는 확률이론과 그래프이론을 결합하여 확률분포(Probability Distribution)를 표현하고, 확률변수(Random Variable)에 대한 확률을 계산할 수 있는 모델이다[11]. 이는 순환관계를 형성할 수 없는 베이저안 네트워크의 단점을 보완한 순환그래프 형태이며, 그래프는 각 변수마다 할당된 노드를 가질 수 있다[10]. 마르코프 네트워크와 1차 논리가 결합한 형태인 마르코프 논리 네트워크는 1차 논리간의 결합을 네트워크 형태로 나타내기 위한 모델이며, 1차 논리형식으로 표현되는 확률모델에 기반을 둔 확률적 추론을 위한 언어라 할 수 있다. 마르코프 논리 네트워크는 변수 간의 관계를 표현하며, 노드는 각 개념을 나타내고, 엣지는 지식을 1차 논리 공식으로 나타내는 네트워크 모델이다[12,13]. 논리 기호로 표시된 공식들에는 제약의 강도를 나타내기 위해 가중치를 부여한다. 일반적인 1차 논리의 진리값은 해당 공식이 참일 때 1, 거짓일 때 0을 갖는다. 그러나 마르코프 논리는 0과 1 사이의 값으로 가중치를 이용해 논리가 얼마나 강한 제약을 갖는지 나타낼 수 있다. 즉, 논리의 신뢰도에 따라 다른 제약을 부여하는 것이며, 논리의 신뢰도가 낮을 경우에는 낮은 제약을 부여하고, 논리를 신뢰할 수 있다면 높은 가중치를 부여하는 것이다. 마르코프 네트워크는 논리와 확률기반의 추론을 제공할 수 있다. 추론에 의한 확률 값은 마르코프 논리 네트워크에서 학습할 수 있으며, 이를 통해 정의하기 어려운 문제를 확률적으로 분류할 수 있다.

## 3. 확률적 추론 모델을 이용한 단문 텍스트 분류

### 3.1 제안한 시스템의 개요

본 논문은 카테고리별 특징을 정의 및 추출하고, 특징을 고려한 가중치 방법을 적용하여, 단문 텍스트 분류에 관한 연구이다. 본 시스템은 학습데이터셋을 구축하는 과정과 학습데이터셋을 기반으로 단문 텍스트를 분류하는 과정으로 구성된다. 다음 Fig. 1은 가중치가 부여된 특징의 학습과 추론규칙의 적용을 통해 단문 텍스트 분류를 수행하기 위한 시스템 구성도를 나타낸다.

### 3.2 학습데이터셋 구축을 위한 키워드 추출

본 장에서는 학습데이터셋의 구축을 위해, 각 카테고리의 키워드를 추출하는 과정을 기술한다.

학습데이터셋을 구축하기 위해서 4가지 카테고리 ('Advertisement', 'Review', 'News', 'Private') 별 트위터 데이터를 수집하고, 카테고리 별 수집된 데이터의 전처리과정을 거쳐 키워드를 추출한 후, 특징정보기반의 키워드가중치(Category Term Weight)를 부여하여 학습데이터셋을 구축한다.

본 논문에서는 학습데이터셋에서 등장하는 동사를 키워드로 정의하였다. 일반적으로 문서의 주제별 분류의 경우, 명사는 문장에서 주어 역할을 하기 때문에 명사를 키워드로 선정하게 되지만, 본 연구에서는 문서의 유형별 분류가 목적이기 때문에 문장에서 내용을 전개하는 서술어 역할을 하는 동사를 키워드로 선정하였다. 키워드를 추출하기 위한 전처리과정은 다음과 같다. 첫 번째 단계에서 특수문자와 숫자

를 제거하고, 두 번째 단계에서 문장 내 의미 판별에 큰 영향을 미치지 않는 관사, 전치사, 조사, 접속사 등의 불용어를 제거한다. 세 번째 단계에서 POS (Part-Of-Speech) 태깅을 통해 동사만을 추출하였으며, 키워드 빈도수(Term Frequency) 측정과 키워드가중치 계산을 위해 스테밍(Stemming)을 거쳐 추출된 동사의 원형을 Term\_List DB에 저장한다. 다음 Table 1은 키워드를 추출하는 과정을 나타낸다.

앞선 과정을 통해 추출된 키워드는 카테고리 내에서의 키워드 빈도수와 Index no.와 함께 'Term\_List DB'에 저장된다. 다음 Table 2는 Term\_List DB 구조를 나타낸다.

### 3.3 단문 텍스트 분류를 위한 특징정보 기반의 가중치 측정

모든 문서는 주제를 가지고 있지만 문서에서 어떤 글의 주제는 직접적으로 관찰 할 수가 없다. 그래서 모든 텍스트는 여러 가지 주제가 혼합되어 확률적으로 중점된 상태라고 할 수 있으며, 각 단어들은 이런 주제들 중 하나를 반영하고 있다. 그러므로 각 단어가 나타내는 주제들을 확률적인 값으로 나타 낼 수 있다. 본 논문과 같이 자동 문서분류를 위한 방법으로 확률적 방법을 사용할 경우 분류의 정확성을 높이기 위해 추출된 키워드에 대한 가중치를 부여한다. 이는 한 문서가 가지고 있는 개념의 주제적 요소로서의 중요도에 따라 상대적 가치를 표현하기 위함이다.

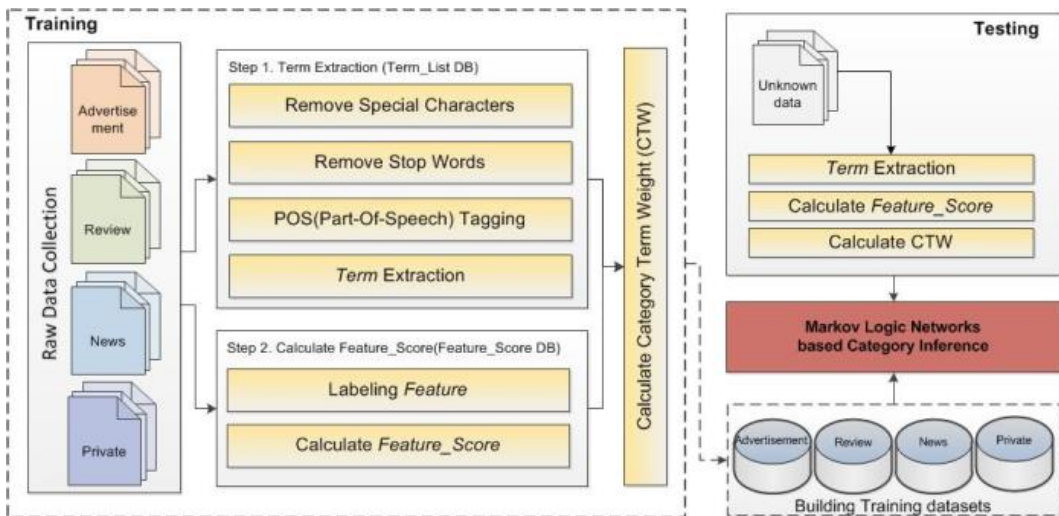


Fig. 1. Architecture of the Proposed System.

Table 1. Example of the term extraction process

Step	Description	
	Index no.	Sample of Data in 'advertisement'
Raw data of twitter	001	#HappyPrimeDay deal: Save up to <u>50%</u> off select Philips headphones @Amazon <a href="http://amzn.to/1SnVu6I">http://amzn.to/1SnVu6I</a>
	002	Traveling this weekend? Save up to <u>\$2000</u> on select travel accessories
	200	Olivia Jordan of Oklahoma was crowned Miss USA
	300	I bought a few things at ikea to make my room travel themed, inspired by jc
Preprocessing	001	HappyPrimeDay deal Save select Philips headphones
	002	Traveling weekend Save select travel accessories
	200	Olivia Jordan Oklahoma crowned Miss USA
	300	I bought things ikea make room travel themed inspired
Term Extraction	001	deal save select
	002	save select
	200	crown
	300	buy make inspire

Table 2. The Structure of Term\_List Database

Extracted Term	Term Frequency	Index no.
buy	1	300
deal	1	001
crown	1	200
inspire	1	300
make	1	300
save	2	001,002
select	2	001,002

본 논문의 키워드 가중치는 각 카테고리에서 추출된 키워드의 빈도수를 기반으로 계산된다. 하지만, 출현하는 단어의 빈도수에만 의존할 경우 카테고리의 정

보를 제대로 반영하기 어렵다. 이러한 문제를 극복하기 위해 본 논문에서는 추출된 키워드 빈도수와 특징 정보를 결합하여 특징기반 가중치 측정 방법을 제안하고자한다. 다음 Table 3은 SNS 메시지의 유형을 분석한 선행연구[14]를 바탕으로 정의한 특징정보를 나타낸다. 앞서 추출한 키워드에 카테고리별 가중치를 부여하기 위해 각 카테고리(News, Opinion, Advertisement, Private)의 성향을 나타내는 특징을 정의하였다.

키워드가중치를 구하기 위해서 정의된 특징 정보를 바탕으로 특징 점수(Feature Score)를 계산하여 'Feature\_Score DB'에 저장한다. 특징 점수는 한 개의 메시지에서 나타나는 특징의 개수를 의미한다. 다

Table 3. Feature Details

Feature	Description		Labeling feature
	Index no.	Sample of data	
f4 : (Number%)	001	#HappyPrimeDay deal: Save up to <u>50%</u> off select Philips headphones @Amazon <a href="http://amzn.to/1SnVu6I">http://amzn.to/1SnVu6I</a>	50%(f4)
f5 : (URL)			<a href="http://amzn.to/1SnVu6I">http://amzn.to/1SnVu6I</a> (f5)
f6 : (@ID)			@Amazon(f6)
f3 : (\$Number)	002	Traveling this weekend? Save up to <u>\$2000</u> on select travel accessories	\$2000(f3)
f1 : (Person_name)	200	<u>Olivia Jordan</u> of <u>Oklahoma</u> was crowned Miss <u>USA</u>	Olivia Jordan(f1)
f2 : (Place_name)			Oklahoma(f2), USA(f2)
f7 : (1st-person pronouns)	300	I bought a few things at ikea to make <u>my</u> room travel themed, inspired by jc	I(f7), my(f7)

Table 4. Feature Score Database

Extracted Term	Term Frequency	Feature Score	Category Term Weight
buy	5	2,3,1,2,3	7.2
deal	3	3,1,2	5
crown	1	3	4
inspire	1	2	3
make	1	2	3
save	4	3,1,2,3	6.25
select	5	3,1,2,2,2	7

음 식 (1)는 문서분류를 위한 가중치 계산을 위해 키워드 빈도수와 특징 점수를 결합한 수식이다. 다음 수식을 이용하여 특징정보기반의 키워드가중치를 구하면 다음 Table 4와 같다.

$$CategoryTermWeight(CTW) = TF + \frac{Feature\ Score}{TF} \quad (1)$$

### 3.4 각 카테고리 별 마르코프 논리 네트워크 학습 및 추론

본 장에서는 학습문서로부터 추출한 각 카테고리의 키워드에 가중치를 부여하여 학습데이터를 구축하였고, 마르코프 논리 기반의 카테고리 추론을 위해 학습 데이터들을 1차 논리 규칙들로 표현한다. 일반적인 1차 논리규칙은 참 또는 거짓인 진리 값을 갖지만, 마르코프 모델에서는 규칙들이 가중치 값을 가질 수 있다. 마르코프 논리 네트워크 추론을 위해 사용된 학습 데이터들을 1차 논리로 표현한 예는 다음 Table 5와 같다.

$Term(save)$ 는 ‘Advertisement’ 카테고리에서 추출된 키워드를 의미하고,  $hasCategory(save, Advertisement)$ 는 ‘save’가 ‘Advertisement’ 카테고리에

Table 5. Sample of Training datasets

CTW	Data expression
0.0	Term(Buy)
0.0	Term(Deal)
0.0	Term(Save)
...	...
7.2	hasCategory (Buy,Advertisement)
5.0	hasCategory (Deal,Advertisement)
6.25	hasCategory (Save,Advertisement)
...	...

속해있음을 표현한다. 각 관계는 가중치를 갖는다. 학습과정이 끝나면, 미분류 문서에 대한 카테고리 추론을 위해 학습데이터를 기반으로 추론규칙을 적용한다. 본 논문에서는 Apriori 알고리즘[15]을 바탕으로 문서와 특성을 고려하여 마르코프 논리에 적용할 수 있는 추론 규칙을 설계하였다. Apriori 알고리즘은 데이터의 클래스를 예측하는데 연관 규칙을 이용하는 연관 규칙 분류(Associative Classification)[16] 방법의 하나로써, A 항목 집합이 존재하면 C 라는 클래스에 속하는 경향이 있음을 의미하는 규칙(“A →C”)을 기반으로 문서를 분류하는 이론이다. 다음 Table 6은 본 연구에서 구축한 학습데이터셋을 기반으로 미분류 문서의 카테고리 추론을 위해 설계한 추론 규칙의 예이다.

설계된 규칙을 기반으로 문서의 카테고리를 추론하기 위해 미분류 문서의 특징들을 추출하고, 마르코프 논리 네트워크에서의 정의된 규칙과 가중치들을 바탕으로  $hasCategory$  확률이 높은 카테고리가 미분류 문서의 카테고리로 선정된다.

### 4. 실험 결과 및 고찰

본 연구에서 사용한 카테고리는 ‘Advertisement’, ‘Reviews’, ‘News’, ‘Private’ 4개의 카테고리이며, 기존 단문텍스트 분류에 관한 선행연구[17-21]들을 기반으로 선정하였으며, 단문텍스트를 이용하는 대표

Table 6. Sample of Inference Rules

Inference Rule
//predicate declarations Term(term) Advertisement(term) Review(term) News(term) Private(term) hasCategory(term,category) category = {Advertisement, Review, News, Private}
//Rules Term(x) ^ hasCategory(x,y) Term(x) ^ !hasCategory(x,y) Term(x) ^ hasCategory(x,Advertisement) => Advertisement(x) Term(x) ^ hasCategory(x,Review) => Review(x) Term(x) ^ hasCategory(x,News) => News(x) Term(x) ^ hasCategory(x,Private) => Private(x)

적인 SNS인 트위터로부터 카테고리별 데이터를 수집하였다. 제안 기법의 실험과 성능 평가를 위해 트위터에서 각 카테고리별 메시지를 500건씩 수집하였으며, 400건의 학습데이터와 100건의 실험데이터로 데이터셋을 구축하였다. 그 중 전처리 과정을 거쳐 선정된 키워드에는 본 논문에서 제안한 가중치 부여 방법을 적용하여, 학습 데이터를 구축하였으며, 미분류 문서의 카테고리 추론을 위해 추론 규칙을 적용하였다. 마르코프 논리의 적용은 공개소프트웨어인 Alchemy(<http://alchemy.cs.Washington.edu>)를 활용하여 추론 규칙에 따라 확률적 추론을 수행하였다. 제안한 특징기반 가중치의 성능판단을 위해서 기존의 키워드 빈도수 기반의 문서 분류 방법과 비교 실험을 수행하였다. 다음 Table 7은 키워드 빈도수 기반의 트위터 메시지 분류 결과이다. 'Reviews'의 경우 다양한 분야의 주제에 대한 의견이 등장하기 때문에 추출된 키워드의 빈도수가 높지 않기 때문에 비교적 낮은 성능을 보였다. 다음 Table 8는 CTW 방법을 사용한 트위터 메시지 분류의 결과이다. 일반적인 분류 알고리즘은 학습데이터셋을 기반으로 분류를 수행하기 때문에, 학습데이터가 각 카테고리의 특징을 얼마나 잘 서술하는지가 성능에 많은 영향을 준다. 기존 문서 분류에서는 문서에 등장하는 단어 간의 의미적 또는 통계적 정보를 특징으로 이용하여 학습 데이터셋을 구축하지만, SNS 메시지가 단문 텍스트로 구성되어 문맥정보를 활용하기가 어렵다. 따라서, 본 논문에서는 SNS 메시지 유형별 특징정보를 선정

Table 7. Experimental Results of Term Frequency

	Precision	Recall	F-measure
Advertisement	0.716	0.765	0.740
Reviews	0.314	0.594	0.411
News	0.652	0.543	0.592
Private	0.863	0.577	0.691
Average	0.560	0.630	0.440

Table 8. Experimental Results of Feature Weight

	Precision	Recall	F-measure
Advertisement	0.825	0.821	0.822
Reviews	0.522	0.848	0.646
News	0.642	0.617	0.613
Private	0.896	0.905	0.901
Average	0.560	0.000	0.000

하고, 그를 활용한 가중치를 부여하는 방법을 제안하였으며, 본 논문에서 제안한 방법을 적용하여 단문 텍스트를 분류한 결과 기존의 빈도수 기반의 가중치 기법에 비해 성능이 향상되었음을 확인할 수 있었다.

## 5. 결 론

인터넷 이용환경의 변화로 사용자들은 시간과 장소에 구애받지 않고 인터넷을 이용할 수 있게 되었다. 사용자는 다양한 정보와 개인의 일상을 SNS를 통해 공유할 수 있게 되었다. 따라서 실시간으로 발생하는 데이터 분석에 대한 연구가 이뤄지고 있으며, 감정분석, 데이터 분류 등 그 주제 또한 다양하다. 본 논문에서는 SNS특징정보기반의 키워드 가중치를 활용하여 단문텍스트의 분류하였으며, 마르코프 논리 네트워크기반의 추론규칙을 설계 및 적용하여 카테고리 추론을 수행하였다. 본 논문에서 제시하는 방안은 기존의 빈도수 기반의 분류 방법에 비해 분류 성능을 향상 시킬 수 있음을 실험을 통해 보였으며, 향후 연구로는 문서 분류 성능 향상을 위해 다양한 특징들을 선정하고, 다양한 추론 규칙을 고려한 한 연구를 수행하고자 한다.

## REFERENCE

- [1] Wikipedia, [http://en.wikipedia.org/wiki/Information\\_overload](http://en.wikipedia.org/wiki/Information_overload) (accessed Mar., 07, 2017).
- [2] M. Nam, E.. Lee, and J. Shin, "A Method for User Sentiment Classification using Instagram Hashtags," *Journal of Korea Multimedia Society*, Vol. 18, No. 11, pp. 1391-1399, 2015.
- [3] B. Ko, D. Choi, C. Choi, J. Choi, and P. Kim, "Data Classification through Specified Building n-gram," *Proceedings of the International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, pp. 171-176, 2012.
- [4] T.M. Mitchell, *Machine Learning*, McGraw-Hill Science/Engineering/Math, 1997. New York.
- [5] B. Ko, K. Oh, and P. Kim, "A Research for Web Documents Genre Classification Using

- STW,” *Journal of Information Technology and Architecture*, Vol. 9, No. 4, pp. 413-422, 2012.
- [6] Wikipedia, <http://en.wikipedia.org/wiki/Tf%E2%80%93idf> (accessed Mar., 07, 2017).
- [7] Wikipedia, [http://en.wikipedia.org/wiki/Mutual\\_information](http://en.wikipedia.org/wiki/Mutual_information) (accessed Mar., 24, 2017).
- [8] C. Han, S. Park, and S. Lee, “A Document Classification System Using Modified ECCD and Category Weight for each Document,” *Korea Information Processing Society*, Vol. 19B, No. 4, pp. 237-242, 2012.
- [9] M. Richardson and P. Domingos, “Markov logic networks,” *Journal Machine Learning*, Vol. 62, No. 1-2, pp. 107-136, 2006.
- [10] S. Riedel and M.R. Ivan, “Collective Semantic Role Labelling with Markov Logic,” *Proceedings of the international Conference on Computational Natural Language Learning*, pp. 193-197, 2008.
- [11] C. Choi, J. Choi, E. Lee, I. You, and P. Kim, “Probabilistic Spatio-temporal Inference for Motion Event Understanding,” *Neurocomputing*, Vol. 122, pp. 24-32, 2013.
- [12] P. Oliveira, *Probabilistic Reasoning in the Semantic Web using Markov Logic*, Master’s Thesis of University of Coimbra, 2009.
- [13] P. Domingos and D. Lowd, *Markov Logic: An Interface Layer for Artificial Intelligence*, Morgan and Claypool Publishers, San Francisco, California, 2009.
- [14] G. Song, Y. Ye, X. Du, X. Huang, and S. Bie, “Short Text Classification : A Survey,” *Journal of Multimedia*, Vol. 9, No. 5, pp. 635-643, 2014.
- [15] L. Meng, R. Huang, and J. Gu, “A Review of Semantic Similarity Measures in WordNet,” *International Journal of Hybrid Information Technology*, Vol. 6, No. 1 pp. 1-12, 2013.
- [16] B. Liu, W. Hsu, and Y. Ma, “Integrating Classification and Association Rule Mining,” *Proceedings of Knowledge Discovery and Data Mining*, pp. 80-86, 1998.
- [17] B. Siram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, “Short Text Classification in Twitter to Improve Information Filtering,” *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 841-842, 2010.
- [18] M. Tare, I. Gohokar, J. Sable, D. Paratwar, and R. Wajgi, “Multi-Class Tweet Categorization using Map Reduce Paradigm,” *International Journal of Computer Trends and Technology*, Vol. 9, No. 2, pp. 78-81, 2014.
- [19] I. Dilrukshi and K. Zoysa, “A Feature Selection Method for Twitter News Classification,” *International Journal of Machine Learning and Computing*, Vol. 4, No. 4, pp. 365-370, 2014.
- [20] J. Wang, G. Cong, W. Zhao, and X. Li, “Mining User Intents in Twitter : Semi-Supervised Approach to Inferring Intent Categories for Tweets,” *Proceeding of 29th Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence*, pp. 339-345, 2015.
- [21] Y. Chun, “A SNS Message Type Classification System using Language Independent Features and Dependent Features,” *International Journal of Software Engineering and Its Applications*, Vol. 9, No. 7, pp. 151-158, 2015.



이 은 지

2012년 조선대학교 컴퓨터공학부  
졸업(공학사)  
2015년 조선대학교 컴퓨터공학과  
박사수료  
2017년 현재 조선대학교 석박사  
통합과정

관심분야 : 자연어처리, 데이터 마이닝, 시맨틱 웹, 온톨로지, 지능형 추론 기술 등



김 판 구

1988년 조선대학교 컴퓨터공학과  
졸업(공학사)  
1990년 서울대학교 컴퓨터공학과  
석사졸업(공학석사)  
1994년 서울대학교 컴퓨터공학과  
박사졸업(공학박사)

관심분야 : 지능형 정보처리, 시맨틱 웹, 온톨로지, 자연어처리, 데이터 마이닝 등